



Universidade de Brasília - UnB  
Instituto de Ciências Exatas - IE  
Departamento de Estatística - EST

# **Segmentação por Modelos de Classe Latente: Uma aplicação ao corpo discente da Universidade de Brasília**

Arthur Soares de Albuquerque

Orientador: Professora Ana Maria Nogales Vasconcelos

Brasília

2017



Arthur Soares de Albuquerque

**Segmentação por Modelos de Classe Latente: Uma  
aplicação ao corpo discente da Universidade de Brasília**

Relatório parcial apresentado à disciplina de Trabalho de Conclusão de Curso I de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Professora Ana Maria Nogales Vasconcelos

Brasília

2017

Monografia intitulada “Segmentação por Modelos de Classe Latente: Uma aplicação ao corpo discente da Universidade de Brasília”, defendida por Arthur Soares de Albuquerque que será apresentada no primeiro semestre de 2018, na Universidade de Brasília - Distrito Federal, pela banca examinadora assim constituída:

---

**Professora Ana Maria Nogales  
Vasconcelos**  
Orientador

---

**Professora Maria Teresa Leão Costa**  
Membro da Banca

---

**Professor Luís Gustavo do Amaral  
Vinha**  
Membro da Banca

Brasília  
2017

# Lista de tabelas

|  |    |
|--|----|
| Tabela 1 – Classificação dos Métodos Utilizados para Análise de Segmentação . .                                    | 10 |
| Tabela 2 – Renda média domiciliar em cada classe socioeconômica segundo o<br>Critério Brasil - ABEP 2015 . . . . . | 18 |
| Tabela 3 – Critérios de Seleção para Modelos de Classe Latente com S segmentos .                                   | 23 |
| Tabela 4 – Parâmetros Estimados no Modelo de Classe Latente com 9 Segmentos  | 23 |
| Tabela 5 – Frequências absolutas e relativas por segmentos . . . . .   | 25 |
| Tabela 6 – Frequências absolutas e relativas por segmentos em cursos de alto prestígio                             | 26 |
| Tabela 7 – Frequências absolutas e relativas por segmentos em cursos de médio<br>prestígio . . . . .               | 27 |
| Tabela 8 – Frequências absolutas e relativas por segmentos em cursos de baixo<br>prestígio . . . . .               | 27 |



# Sumário

|            |   |           |
|------------|---|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>   | <b>7</b>  |
| <b>2</b>   | <b>APLICAÇÃO EM DADOS EDUCACIONAIS</b>  | <b>8</b>  |
| <b>2.1</b> | <b>O Sistema Educacional</b>  | <b>8</b>  |
| <b>3</b>   | <b>OBJETIVOS</b>  | <b>9</b>  |
| <b>3.1</b> | <b>Objetivo Geral</b>   | <b>9</b>  |
| 3.1.1      | Objetivo Específico   | 9         |
| <b>4</b>   | <b>A TÉCNICA DE SEGMENTAÇÃO</b>   | <b>10</b> |
| <b>4.1</b> | <b>Etapas para a realização de uma Análise de Segmentação</b>                       | <b>10</b> |
| <b>5</b>   | <b>SEGMENTAÇÃO POR MODELOS DE CLASSE LATENTE OU MISTURA FINITA</b>                  | <b>11</b> |
| <b>5.1</b> | <b>Introdução aos Modelos de Classe Latente ou Mistura Finita</b>                   | <b>11</b> |
| <b>5.2</b> | <b>Metodologia do Modelo</b>  | <b>12</b> |
| <b>5.3</b> | <b>Estimação dos Parâmetros</b>   | <b>13</b> |
| 5.3.1      | O Algoritmo EM  | 13        |
| <b>5.4</b> | <b>CrITÉrios para seleção do modelo</b>   | <b>14</b> |
| 5.4.1      | CrITÉrio de Informação de Akaike (AIC)  | 14        |
| 5.4.2      | CrITÉrio de Informação de Bayes (BIC)   | 14        |
| <b>5.5</b> | <b>Testes de adequabilidade</b>   | <b>14</b> |
| 5.5.1      | Qui-Quadrado de Pearson Goodness of Fit   | 15        |
| 5.5.2      | Teste Qui-Quadrado de Razão de Verossimilhança ou Deviance ( $G^2$ )                | 15        |
| <b>6</b>   | <b>UMA APLICAÇÃO AOS ESTUDANTES DA UNIVERSIDADE DE BRASÍLIA</b>                     | <b>16</b> |
| <b>6.1</b> | <b>Os dados do do Observatório da Vida Estudantil</b>                               | <b>16</b> |
| <b>6.2</b> | <b>O Questionário</b>   | <b>16</b> |
| <b>6.3</b> | <b>As Variáveis de Estudo</b>   | <b>16</b> |
| 6.3.1      | Sistema de ingresso na Universidade   | 17        |
| 6.3.2      | CrITÉrio Brasil   | 17        |
| 6.3.3      | PrestÍgio do curso  | 18        |
| <b>7</b>   | <b>ANÁLISE DE DADOS</b>   | <b>19</b> |
| <b>7.1</b> | <b>Análise socioeconômica descritiva dos estudantes ingressantes na UnB em 2017</b> | <b>19</b> |

|     |  |           |
|-----|--|-----------|
| 7.2 | Identificando os Segmentos . . . . .           | 22        |
| 8   | CONCLUSÃO . . . . .                            | 29        |
|     | REFERÊNCIAS . . . . .                          | 31        |
|     | <b>ANEXOS</b>                                  | <b>33</b> |
|     | ANEXO A – SINTAXE DE PROGRAMAÇÃO SAS . . . . . | 34        |



# 1 Introdução

Define-se como segmentação uma forma de separação de indivíduos em categorias diferenciadas. Amplamente desenvolvido na área de marketing, a análise de segmentação surgiu com o intuito de entender melhor o mercado e identificar as necessidades e desejos dos consumidores, redefinir estratégias de mercado e aumentar o lucro. Essa técnica, ao identificar o número de segmentos e o tamanho de cada um, faz uma distinção que, ao dividir um mercado com demanda heterogênea em sub mercados com demanda homogênea, viabiliza um ajuste mais adequado de marcas, produtos ou serviços às necessidades dos consumidores permitindo estabelecer conjuntos que compartilham perfis semelhantes. Isto posto, como resultado, empresas podem ter um maior entendimento do comportamento do consumidor, possibilitando a criação de ofertas dirigidas a um determinado segmento, o desenvolvimento de novos produtos ou serviços e, também, a identificação de interesses para proporcionar uma melhor comunicação com o cliente.

Segundo Wedel & Kamakura (2000), bens não podem mais ser produzidos e vendidos sem considerar as necessidades dos clientes e reconhecer a heterogeneidade dessas necessidades. Deste modo, a análise de segmentação aparece como um método para a realização de uma análise estratégica mais cuidadosa em que, ao identificar segmentos, se torna um meio mais eficaz para a seleção de mercados com menor dimensão porém mais homogêneos em suas preferências e comportamentos observados.

Exemplos de aplicação dos métodos de segmentação podem ser encontrados no mercado empresarial, segmentação para otimização de serviços de qualidade, segmentação para avaliação de satisfação do consumidor, segmentação demográfica, segmentação no setor bancário, entre outros. Para este trabalho, propõe-se uma aplicação na área da educação com o objetivo de trazer contribuições acadêmicas e sociais para a realidade atual do país.

## 2 Aplicação em dados educacionais

### 2.1 O Sistema Educacional

O sistema educacional possui importância essencial no desenvolvimento socioeconômico de um país. E com isso, a produção de dados e informações estatístico-educacionais de forma ágil e fidedigna, que retrate a realidade do setor educacional, é o instrumento básico de avaliação, planejamento e auxílio ao processo decisório para o estabelecimento de políticas de melhoria da educação brasileira. (Castro, 2000).

Para isso, torna-se, então, cada vez mais necessário, levantar dados, avaliar e monitorar programas existentes, entender como o processo de aprendizagem ocorre e identificar os fatores que influenciam esse processo a fim de subsidiar discussões e ações educativas (Karino, 2016). Portanto, avaliou-se que seria de grande importância a realização de uma investigação sobre o corpo discente da Universidade de Brasília de modo a acompanhar e entender melhor esse cenário.

Segundo Silva, Barbosa, Sousa (2006), a universidade brasileira está marcada por grandes desigualdades sociais. Neste sentido, dependendo da classe social do estudante, não apenas o acesso, mas também a permanência no nível superior de ensino apresentam obstáculos para se concretizar. Para isso, políticas públicas como o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), e o Sistema de Cotas Sociais e Raciais (Lei no 12.711/2012) foram criadas com a intenção de reduzir as desigualdades no acesso e na permanência de grupos sociais menos privilegiados na universidade.

Portanto, diante de um cenário de transformações, para uma melhor identificação e compreensão dos estudantes da Universidade de Brasília, propõe-se utilizar uma técnica de segmentação como ferramenta para se obter tipologias ao utilizar a heterogeneidade do corpo discente para a obtenção de diferentes perfis homogêneos entre si, caracterizando o estudante. Para isso, serão utilizados os dados obtidos através do questionário, aplicado durante o registro, disponibilizados pelo Observatório da Vida Estudantil da Universidade de Brasília (OVE).

## 3 Objetivos

### 3.1 Objetivo Geral

Aplicar a técnica e gerar segmentos de modo a identificar perfis dos estudantes de graduação da Universidade de Brasília

#### 3.1.1 Objetivo Específico

- Implementar a técnica de segmentação utilizando o software SAS
- Encontrar o número adequado de perfis por meio de ferramentas estatísticas
- Obter perfis através de uma aplicação do método de segmentação por modelagem de classe latente
- Analisar os segmentos obtidos na aplicação e interpretar os resultados

## 4 A Técnica de Segmentação

### 4.1 Etapas para a realização de uma Análise de Segmentação

O processo para a realização de uma análise de segmentação pode ser dividido em três categorias de técnicas analíticas como etapas do processo. Definidas como: preparação de dados, análise de dados e classificação.

Na etapa de preparação dos dados, técnicas para um melhor gerenciamento ou redução das variáveis podem ser usadas para auxiliar o processo de segmentação dependendo do objetivo. Para essa etapa, destaca-se a análise fatorial, a análise de correspondência, a análise conjunta, entre outras técnicas utilizadas.

Para a etapa de análise de dados, é essencial que a escolha da técnica seja adequada. Para isso, os métodos de segmentação podem ser classificados de acordo com dois aspectos. O primeiro aspecto refere-se à determinação do tipo e número de segmentos. Segundo Wind (1978), a segmentação classifica-se em a-priori e post-hoc. A segmentação a-priori acontece quando o tipo e o número de segmentos são determinados com antecedência pelo pesquisador, enquanto a post-hoc acontece quando o tipo e o número de segmentos são determinados com base nos resultados da análise de dados. O segundo aspecto relaciona-se aos métodos estatísticos utilizados, classificando em métodos descritivos e preditivos. De acordo com Wedel & Kamakura (2000), os métodos descritivos analisam associações entre um só conjunto de base de segmentação, sem distinção entre variáveis dependentes e independentes. Enquanto os preditivos analisam a associação entre dois conjuntos de variáveis em que um conjunto, representado por variáveis dependentes, é explicado por um outro conjunto de variáveis independentes.

Tabela 1 – Classificação dos Métodos Utilizados para Análise de Segmentação

|                   | <b>A priori</b>  | <b>Post hoc</b>   |
|-------------------|--|---|
| <b>Descritiva</b> | Tabelas de contingência,<br>Modelos log-lineares                     | Métodos por Cluster, Sobreposição,<br>Técnicas Fuzzy, ANN, Modelos de Mistura |
| <b>Preditiva</b>  | Tabelas Cruzadas, Regressão,<br>Logística e Análise de Discriminante | AID, ANN, CART,<br>Modelos de Mistura, Regressão Clusterwise                  |

A etapa de classificação tem como finalidade a criação de sistemas de classificação para os segmentos. Sendo assim, usa-se meios para identificar potenciais componentes de cada segmento com base em informações limitadas (podendo ser obtidos de fontes secundárias). Dentre as técnicas utilizadas encontra-se a análise discriminante, regressão múltipla, logística, entre outras.

## 5 Segmentação por Modelos de Classe Latente ou Mistura Finita

### 5.1 Introdução aos Modelos de Classe Latente ou Mistura Finita

Neste trabalho, a técnica de segmentação a ser utilizada será por meio de um modelos de mistura finita (também chamado de classe latente quando as distribuições associadas a cada classe, segmento, são do tipo discreto), considerado o método mais eficiente de segmentação para o estudo de estruturas de preferências.

A análise de classe latente é uma técnica estatística para a análise de dados categóricos multivariados. Muito utilizada quando deseja-se investigar a relação entre as variáveis observadas, identificar e caracterizar clusters, e aproximar a distribuição das observações através das muitas variáveis de interesse. Como em pesquisas de opinião pública, dados de votação individual, comportamento do consumidor, entre outros exemplos.

Define-se como modelos de mistura aqueles em que não é conhecido o segmento no qual pertence o indivíduo ou quantos são dispostos inicialmente. Desta maneira, assume-se que os indivíduos provêm de diversos segmentos homogêneos misturados em proporções desconhecidas. Assim sendo, apresentam como objetivo se desfazer dessas misturas de modo a ser revelado o número de segmentos e a definição das suas estruturas que compõem a população em análise (Calado, 2008). Para cada indivíduo, esses modelos estimam a probabilidade de pertencimento a cada segmento obtido, de modo a perceber relações que o indivíduo possui com os grupos por meio de uma estrutura “Fuzzy”.

Esses modelos surgiram como uma exceção ao uso continuado de técnicas de análise de cluster, e revelaram ser procedimentos mais eficazes para lidar com a heterogeneidade dos parâmetros estudados. Embora sejam considerados generalizações do método k-médias, os segmentos são constituídos de forma probabilística e não determinística como no modelo k-médias. Além disso, são extremamente flexíveis quanto ao tipo de variáveis ativas e precisos também em bases de dados relativamente pequenas, portanto, dificilmente haveria problemas de adequabilidade da técnica.

Ao passo que os modelos tradicionais usados em estudos para classificar tipologias, tais como a análise de regressão, discriminante e log-linear contêm parâmetros que descrevem apenas as relações entre as variáveis observadas. Os modelos de classe latente (ou mistura finita) são diferentes, pois incluem uma ou mais variáveis discretas não observadas que podem ser utilizadas para explicar a associação existente entre as variáveis observadas. No contexto deste estudo, considera-se como classes latentes, as categorias dessas variáveis latentes, ou seja, os clusters ou segmentos.

No geral, existem três razões importantes para considerar esse tipo de modelagem. Primeiramente, o grande avanço no desenvolvimento das mais avançadas técnicas de segmentação aconteceram devido ao avanço na teoria e estimativa de modelos de mistura finita. Em segundo, como foi introduzido brevemente, esses modelos fornecem uma estrutura mais flexível para segmentação. E em terceiro, porque ao longo das últimas décadas, os modelos de classes latentes tornaram-se mais acessíveis de serem realizados computacionalmente, ganhando protagonismo na análise de segmentação frente aos métodos tradicionais de grupamento.

## 5.2 Metodologia do Modelo

A mistura consiste em assumir que de um grupo de dados  $Y = (y_1, \dots, y_n)$ , com  $i = 1, \dots, N$ ,  $y_i$  provém de um possível segmento  $s$  não observável (componente ou classe latente) independente e normalmente distribuído,  $s_1, s_2, \dots, S$ , cada um seguindo uma distribuição paramétrica, pertencendo à  $j$  – ésima variável, em que  $j = 1, \dots, J$  e à  $k$  – ésima categoria, com  $k = 1, \dots, K_j$ .

Então pode-se escrever a função distribuição de mistura da seguinte forma:

$$f(y|\pi\theta) = \sum_s \pi_s f_s(y|\theta)$$

Como modelo de classe latente é geralmente utilizado para analisar dados categorizados, assume-se que as observações aleatórias seguem uma distribuição que é uma mistura de distribuições multinomiais multivariadas:

$$f(y|\theta) = \prod_{j=1}^J \prod_{k=1}^{K_j} (\theta_{jsk})^{Y_{ijk}}$$

Portanto, a função densidade de probabilidade para o modelo de classe latente é:

$$f(y|\pi, \theta) = \sum_{s=1}^S \pi_s \prod_{j=1}^J \prod_{k=1}^{K_j} (\theta_{jsk})^{Y_{ijk}}$$

Em que  $\theta_{jsk}$  dá a probabilidade de pertencimento da categoria  $k$ , na variável  $j$  no segmento  $s$ , com  $\sum_{k=1}^{K_j} \theta_{jsk} = 1$ . E  $\pi_s$  a proporção da mistura que, com  $\sum_{s=1}^S \pi_s = 1$ , fornece a probabilidade *a-priori* do segmento  $s$ .

Já a variável latente  $Y_{ijk}$  verifica se o indivíduo  $i$  pertence ao segmento  $s$ . Logo,

$$Y_{ijk} = \begin{cases} 1, & \text{se } i \in S \\ 0, & \text{caso contrário} \end{cases}$$

assumindo ser independente e multinomialmente distribuída.

Então, o modelo de classe latente é um modelo de independência condicional e a

relação entre as variáveis observadas pode ser explicadas pelos segmentos (classes latentes).

## 5.3 Estimação dos Parâmetros

Considerando  $\Phi = (\pi, \theta)$  o vetor de parâmetros do modelo a serem estimados por meio da maximização da função log-verossimilhança.

$$\ln L = \sum_{i=1}^N \ln \sum_{s=1}^S \pi_s \prod_{j=1}^J \prod_{k=1}^{K_j} (\theta_{jsk})^{Y_{ijk}}$$

Essa maximização pode ser obtida através de métodos como o de Máxima Verossimilhança ou Esperança-Maximização (EM), os mais indicados para casos não bayesianos. Para modelos de classe latente, tem-se como padrão o ajuste via o algoritmo EM.

Obtido os estimadores de  $\Phi$ , por meio da fórmula de Bayes, encontra-se as estimativas da probabilidade posterior que cada indivíduo pertence a cada segmento  $s$ .

$$P(s|y_i, \theta_s) = \frac{\pi_s f_s(y_i|\theta_s)}{\sum_{s=1}^S \pi_s f_s(y_i|\theta_s)}$$

### 5.3.1 O Algoritmo EM

Proposto por Dempster et al. (1977), o algoritmo Esperança-Maximização (EM) é um método iterativo para encontrar a máxima verossimilhança ou posterior dos parâmetros dos modelos.

Como em qualquer modelo de mistura finita, o algoritmo EM é aplicável, porque o grau de pertencimento da classe de cada indivíduo é desconhecida e pode ser tratada como dados faltantes (McLachlan and Krishnan, 1997; McLachlan and Peel, 2000).

Neste método, cada iteração consiste em dois passos, um passo E e um passo M:

Passo E: A esperança da verossimilhança é calculada em relação às informações não observadas, dado as observadas, estimando as probabilidades da classe posterior para cada observação.

$$\hat{p}_s = P(s|y_i, \theta_s)$$

Passo M: a esperança da log-verossimilhança é maximizada para cada componente separadamente, em relação ao peso, em ordem ao vetor de parâmetros  $\Phi$ . Então:

$$\hat{\pi}_s = \frac{1}{N} \sum_{i=1}^N \hat{p}_s$$

Os passos E e M são repetidos até a melhora da verossimilhança de modo que o algoritmo atinja um valor especificado de iterações ou um numero máximo.

## 5.4 Critérios para seleção do modelo

Ao contrário de outras técnicas estatísticas que utilizam meios mais exploratórios para se obter dados agrupados, o modelo de classe latente segue principalmente razões teóricas para a seleção dos segmentos. Para isso, conta com uma variedade de ferramentas estatísticas disponíveis para a avaliação do modelo e determinação do número apropriado de classes latentes, como os critérios de informação AIC (Akaike's Information Criteria) (Akaike (1973)) e BIC (Bayes' Information Criteria) (Schwartz, 1978) para tomada de decisões.

Essas medidas procuram encontrar um equilíbrio entre um sub ajustamento e um super ajustamento do modelo de modo que total de parametros do modelo tenha influência na métrica. Portanto, quanto maior for o numero de parametros, maior sera uma penalizacao introduzida. Desta forma, o modelo mais adequado é aquele com o segmento selecionado  $s$  que minimiza os valores dos critérios de informação.

### 5.4.1 Critério de Informação de Akaike (AIC)

Expresso pela fórmula:

$$AIC = 2\ln L + 2\Phi$$

Em que  $\ln L$  é a log verossimilhança do modelo e  $\Phi$  o número de parâmetros estimados.

### 5.4.2 Critério de Informação de Bayes (BIC)

Expresso pela fórmula:

$$BIC = 2\ln L + N$$

Em que  $\ln L$  é a log verossimilhança do modelo,  $\Phi$  o número de parâmetros estimados e  $N$  o tamanho da amostra.

## 5.5 Testes de adequabilidade

Testes de adequabilidade tem a finalidade de avaliar a qualidade do ajuste de modo a verificar o quão bem os dados observados correspondem ao modelo ajustado. Para isso, usa-se métodos como o Qui-Quadrado de Pearson Goodness of Fit e o Teste Qui-Quadrado da Razão de Verossimilhança ou Deviance ( $G^2$ ) para determinar o quão bom o modelo se ajusta aos dados.



Isto posto, são considerados mais adequados os modelos que minimizam o  $\chi^2$  e o  $G^2$ . Vale ressaltar que esses testes não são válidos em casos de poucas observações nas células (mais de 10-20% das células contendo menos de 5 observações).

### 5.5.1 Qui-Quadrado de Pearson Goodness of Fit

Considerando  $\hat{E}_c$ , com  $c = 1, \dots, C$ , que representa o número esperado de casos em cada célula sob um determinado modelo.

$$\hat{E}_c = N \sum_{s=1}^S \hat{\pi}_s \prod_{j=1}^J \theta_{jsk}$$

Logo o teste pode ser calculado como:

$$\chi^2 = \sum_c \frac{(O_c - \hat{E}_c)^2}{\hat{E}_c}$$

Em que  $O_c$  é o número observado de casos em cada uma das células  $C = \prod K_j$  obtidas.

### 5.5.2 Teste Qui-Quadrado de Razão de Verossimilhança ou Deviance ( $G^2$ )

O teste  $G^2$  pode ser calculado da seguinte forma:

$$G^2 = 2 \sum_c O_c \ln\left(\frac{O_c}{\hat{E}_c}\right)$$

Em que  $\hat{E}_c$ , com  $c = 1, \dots, C$ , representa o número esperado de casos em cada célula sob um determinado modelo e  $O_c$  o número observado de casos em cada uma das células  $C = \prod K_j$  obtidas.

## 6 Uma Aplicação aos estudantes da Universidade de Brasília

### 6.1 Os dados do do Observatório da Vida Estudantil

Conforme mencionado anteriormente, os dados utilizados neste trabalho provêm do Observatório da Vida Estudantil da Universidade de Brasília (OVE). Criado em 2010, integrado ao Núcleo de Estudo e Pesquisa em Ensino Superior da UnB (Nesub) e ao Centro de Estudos Avancados Multidisciplinares (CEAM), com o apoio do Decanato de Ensino de Graduação (DEG), o OVE tem como principal propósito a análise do perfil dos estudantes da UnB e a obtenção de conhecimento mais consolidado a respeito do corpo discente da universidade de modo a obter subsídios e contribuir para uma melhora da política acadêmica e de assistência estudantil.

A base do OVE é composta por dados dos estudantes ingressantes a partir do 1º semestre de 2012 até o presente momento. Essas informações objetivam traçar suas origens geográficas, sociais e econômicas, suas trajetórias escolares e profissionais, e suas expectativas e motivações em relação ao ingresso no ensino superior. Para este trabalho, serão avaliados os seguintes períodos: 1º/2012, 2º/2012, 1º/2013, 2º/2013, 1º/2014, 2º/2014, 1º 2015, 2º/2015, 1º/2016 e 2º/2016.

### 6.2 O Questionário

O questionário possui um total de 60 perguntas, e procura avaliar o perfil socioeconômico e demográfico, a trajetória pré-universitária, a inserção universitária, trabalho e perspectivas futuras. Preservando o sigilo e a anonimidade dos estudantes.

Obtido como parte do processo de registro dos estudantes ingressantes na UnB, o preenchimento completo do questionário não é obrigatório, logo os resultados não retratam o conjunto total de estudantes registrados na universidade. Porém, o número de respondentes esteve sempre acima de 85% do total de ingressantes em cada período.

### 6.3 As Variáveis de Estudo

De forma a evidenciar o perfil do estudante de graduação da Universidade de Brasília, considerou-se neste estudo as seguintes variáveis socioeconômicas:

### 6.3.1 Sistema de ingresso na Universidade

- **Sexo:** Masculino / Feminino;
- **Raça/Cor da Pele:** Brancos e Amarelos / Pretos e Pardos / Indígenas;
- **Sistema de ingresso na Universidade:** Cotas raciais / Cotas Sociais / Sistema Universal;
- **Classe Socioeconômica (Critério Brasil):** A (45 a 100 pontos) / B (29 a 44 pontos) / C (17 a 28 pontos) / D (0 a 16 pontos);
- **Escolaridade do pai:** Até o ensino fundamental / Ensino médio completo / Ensino superior completo;
- **Escolaridade da mãe:** Até o ensino fundamental / Ensino médio completo / Ensino superior completo;
- **Categoria administrativa da instituição de ensino médio:** Escola pública / Escola Particular;
- **Classificação dos cursos de acordo com o prestígio:** Alto / Médio / Baixo;
- **Renda mensal da família:** até 3 SM / 3 a 10 SM / 10 a 20 SM / 20 ou mais SM;
- **Realização de curso preparatório:** Sim / Não;

### 6.3.2 Critério Brasil

Conforme a metodologia descrita em Kamakura & Mazzon (2013), o Critério Brasil da Associação Brasileira de Empresas de Pesquisa (ABEP) surgiu como forma de classificação de domicílios. Em 2015, de acordo com Pilli (2015), um novo critério entrou em vigor como regra nas pesquisas de mercado no Brasil e, ao ser incorporado fatores que determinam também o comportamento de consumo, passou a ser considerado uma ferramenta importante para medir desigualdades socioeconômicas ao classificar extratos sociais.

Com base nesse critério e, indo de acordo com o estudo de Costa (2015), criou-se a variável “Classe Socioeconômica” utilizando as pontuações, que variam de zero a cem, referentes aos quesitos contidos no questionário do OVE. Logo, definiu-se as classes socioeconômicas dos estudantes como:

Tabela 2 – Renda média domiciliar em cada classe socioeconômica segundo o Critério Brasil - ABEP 2015

| <b>Extrato Socioeconômico</b> | <b>Renda Média Domiciliar em Reais</b> |
|-------------------------------|--|
| A                             | Mais de 20 mil                         |
| B                             | de 4 mil até 20 mil                    |
| C                             | de mil até 4 mil                       |
| D                             | até mil                                |

### 6.3.3 Prestígio do curso

Analisando-se a demanda e a nota de corte dos processos seletivos, criou-se a variável “prestígio” com o intuito de evidenciar as diferenças socioeconômicas entre os estudantes em função do curso. Sendo assim, os cursos foram divididos em:

- Alto prestígio: Medicina, Direito, Relações Internacionais, as Engenharias, Odontologia, Psicologia, Comunicação Social e Ciências Econômicas;
- Baixo prestígio: Biblioteconomia, Pedagogia, Serviço Social, Letras e Matemática;
- Médio prestígio: os demais cursos.

## 7 Análise de Dados

### 7.1 Análise socioeconômica descritiva dos estudantes ingressantes na UnB em 2017

Observando o percentual dos ingressantes da Universidade de Brasília nos cursos de graduação, percebe-se que o “sexo” dos estudantes, segundo o padrão adotado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), divide-se em cerca de 51% para o sexo masculino e 49% para o sexo feminino conforme expresso no gráfico da Figura 1.

Já o percentual do ingressantes com relação à raça/cor da pele se divide em 52.8% para pretos e pardos, 46.8% para amarelos e brancos, e 0.4% de indígenas. Como demonstrado no gráfico da Figura 2.

Figura 1 – Percentual de ingressantes na UnB segundo sexo, 2017

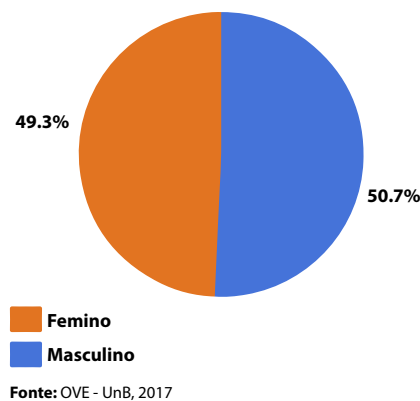
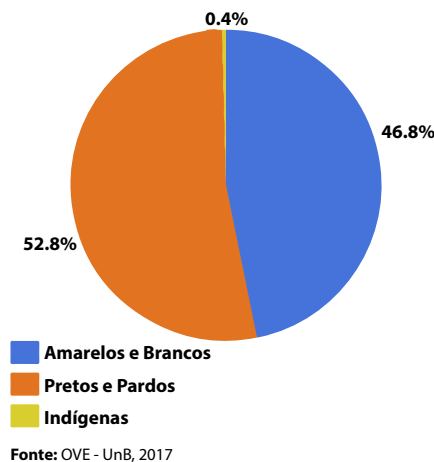
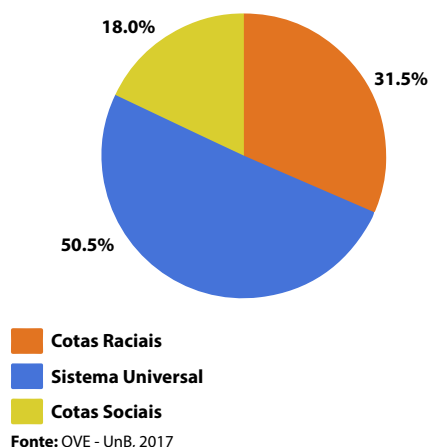


Figura 2 – Percentual de ingressantes na UnB segundo raça/cor da pele, 2017



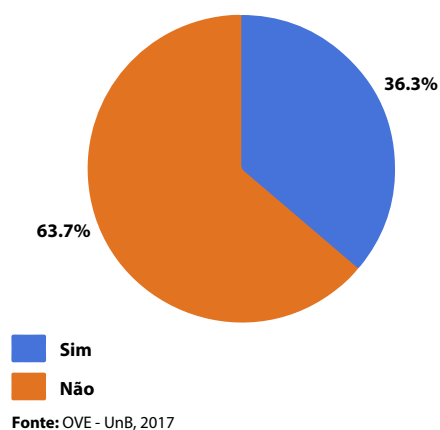
Em relação ao sistema de ingresso do estudante de graduação, por lei (Lei 12.711/12), 50% das vagas são reservadas para estudantes procedentes de escolas públicas. Logo há um equilíbrio em relação aos estudantes provenientes de escolas particulares e um equilíbrio entre os ingressantes por meio do sistema universal e por cotas evidenciado no gráfico da Figura 3.

Figura 3 – Percentual de ingressantes na UnB segundo sistema de ingresso, 2017



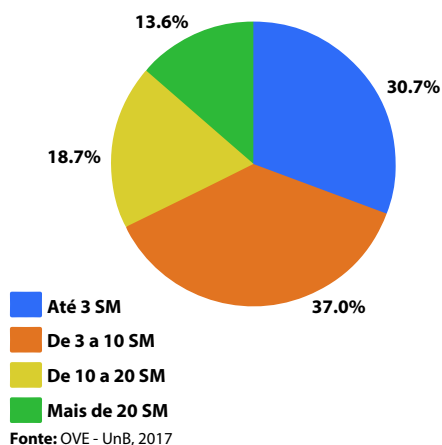
No gráfico da Figura 4, apresenta-se que, dos estudantes de graduação ingressantes, cerca de 36% realizou curso preparatório para entrar na universidade.

Figura 4 – Percentual de ingressantes na UnB segundo realização de curso preparatório, 2017



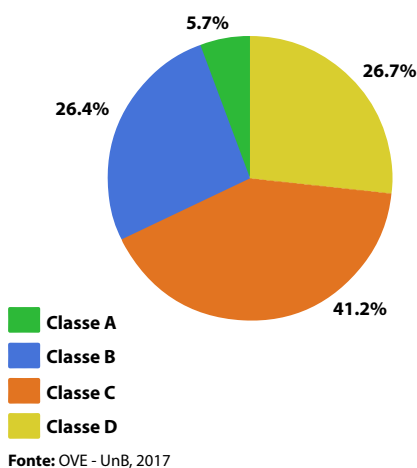
E, pelo gráfico expresso pela Figura 5, verifica-se que a maioria dos estudantes ingressantes se encontram em meio às faixas de renda mais baixas, sendo 37% de renda familiar mensal entre três e dez salários mínimos e 30.7% com renda familiar mensal de até três salários mínimos. Já a minoria dos estudantes ingressantes encontra-se na faixa de renda mais alta, sendo 13.6% com mais de vinte salários mínimos de renda familiar mensal.

Figura 5 – Percentual de ingressantes na UnB segundo renda mensal da família, 2017



Já ao analisar o gráfico da classificação do Critério Brasil, apresentado pela Figura 6, nota-se que a maioria é representada pela classe C, com 41.2%, seguida das classes D e B com porcentagens próximas, 26.7% e 26.4% respectivamente. 5.7% dos estudantes ingressantes foram classificados pertencentes à classe A.

Figura 6 – Percentual de ingressantes na UnB segundo a classificação do Critério Brasil, 2017



Os gráficos enunciados pelas Figuras 7 e 8 se referem, respectivamente, à escolaridade do pai e da mãe do estudante. Em ambos os casos, verificou-se que a maioria dos pais possuem um nível de instrução com no mínimo um ensino superior completo, com 43.6% dos pais e 47.1% das mães. Em seguida, pais e mães com ensino médio completo, 32.2% e 33.8% e, em minoria, pais e mães com baixa escolaridade (até o ensino fundamental) com 24.2% e 19.1% respectivamente.

Figura 7 – Percentual de ingressantes na UnB segundo escolaridade do pai, 2017

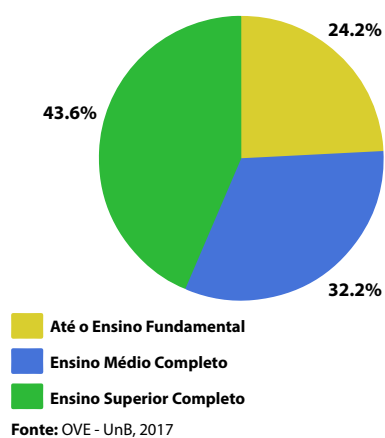
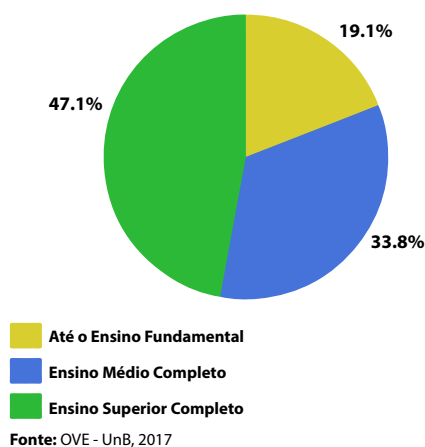


Figura 8 – Percentual de ingressantes na UnB segundo escolaridade da mãe, 2017



## 7.2 Identificando os Segmentos

Em concordância com o que foi discutido na metodologia deste trabalho, o modelo de classe latente procura utilizar diferentes padrões de resposta nos dados para a obtenção dos segmentos. Para isso, deve-se encontrar um número de classes de tal modo que o modelo ainda seja adequado aos dados. Sendo assim, rodou-se o modelo utilizando a PROC LCA no software estatístico SAS e comparou-se as diferentes soluções possíveis utilizando os critérios de seleção, segundo apresentado na Tabela 3.



Tabela 3 – Critérios de Seleção para Modelos de Classe Latente com S segmentos

| S        | lnL              | G2             | AIC            | BIC            |
|----------|------------------|----------------|----------------|----------------|
| 2        | -47442.40        | 7238.43        | 7296.43        | 7497.46        |
| 3        | -46678.22        | 5710.08        | 5798.08        | 6103.10        |
| 4        | -46124.75        | 4603.15        | 4721.15        | 5130.14        |
| 5        | -45056.96        | 2467.56        | 2615.56        | 3128.53        |
| 6        | -44766.79        | 1887.22        | 2065.22        | 2682.17        |
| 7        | -44693.01        | 1739.65        | 1947.65        | 2668.59        |
| 8        | -44646.28        | 1646.19        | 1884.19        | 2709.11        |
| <b>9</b> | <b>-44543.57</b> | <b>1440.78</b> | <b>1708.78</b> | <b>2637.68</b> |

Selecionado o número de segmentos adequado, o modelo tenta conceder grupos que sejam condicionalmente independentes. Logo, dentro de cada segmento, as correlações entre as variáveis serão zero devido ao pertencimento ao determinado grupo.

Tabela 4 – Parâmetros Estimados no Modelo de Classe Latente com 9 Segmentos

| Variáveis                   | Categorias               | Classes Latentes (Segmentos) |         |         |         |         |         |         |         |         |
|-----------------------------|--------------------------|------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                             |                          | 1                            | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       |
| Sistema de Ingresso         | Cotas Raciais            | .042201                      | .002030 | .937645 | .001146 | .862109 | .005091 | .011477 | .000102 | .864663 |
|                             | Sistema Universal        | .107459                      | .997255 | .062182 | .998450 | .039153 | .004465 | .088287 | .999139 | .999139 |
| Sexo                        | Cotas Sociais            | .850341                      | .000715 | .000173 | .000404 | .098738 | .990445 | .900236 | .000760 | .000079 |
|                             | Masculino                | .420593                      | .580941 | .478098 | .513534 | .479151 | .496890 | .556211 | .455472 | .580731 |
| Raça/Cor da Pele            | Feminino                 | .579407                      | .419059 | .521902 | .486466 | .520849 | .503110 | .443789 | .544528 | .419269 |
|                             | Amarelos/Branços         | .735292                      | .764622 | .000478 | .546197 | .018708 | .948085 | .781836 | .580455 | .004128 |
|                             | Pretos/Pardos            | .264704                      | .232283 | .993171 | .453800 | .978780 | .049798 | .208074 | .413744 | .988103 |
| Ensino Médio                | Indígenas                | .000004                      | .003096 | .006351 | .000003 | .002512 | .002117 | .010090 | .005801 | .007769 |
|                             | Escola Pública           | .994368                      | .022910 | .884101 | .533838 | .993026 | .996199 | .996306 | .055606 | .476272 |
| Escolaridade do Pai         | Escola Particular        | .005632                      | .977090 | .115899 | .466162 | .006974 | .003801 | .003694 | .944394 | .523728 |
|                             | Até o Fundamental        | .199419                      | .007372 | .196448 | .650787 | .768842 | .709189 | .068141 | .132092 | .057132 |
| Escolaridade da Mãe         | Ensino Médio Completo    | .621746                      | .102737 | .587945 | .326958 | .221223 | .257008 | .148641 | .465085 | .245667 |
|                             | Ensino Superior Completo | .178834                      | .889891 | .215607 | .022254 | .009935 | .033803 | .783218 | .402824 | .697201 |
| Realizou Curso Preparatório | Até o Ensino Fundamental | .082457                      | .002888 | .110039 | .658235 | .635549 | .694981 | .048640 | .068792 | .025241 |
|                             | Ensino Médio Completo    | .645920                      | .110997 | .614721 | .277047 | .327635 | .302049 | .221008 | .416511 | .173122 |
|                             | Ensino Superior Completo | .271623                      | .886115 | .275240 | .064718 | .036816 | .002970 | .730352 | .514697 | .801637 |
| Renda Mensal da Família     | Sim                      | .354209                      | .458098 | .324986 | .155879 | .233505 | .210160 | .490351 | .359643 | .529660 |
|                             | Não                      | .645791                      | .541902 | .675014 | .844121 | .766495 | .789840 | .509649 | .640357 | .470340 |
|                             | Até 3 SM                 | .480674                      | .003027 | .448362 | .520199 | .787509 | .729569 | .007605 | .168492 | .001481 |
|                             | De 3 a 10 SM             | .489750                      | .174508 | .516546 | .297209 | .189816 | .222487 | .486741 | .610445 | .484802 |
|                             | De 10 a 20               | .027885                      | .370107 | .035016 | .122188 | .015689 | .016308 | .415055 | .213137 | .348883 |
|                             | Mais de 20               | .001691                      | .452358 | .000075 | .060403 | .006986 | .031636 | .090598 | .007927 | .164833 |

De acordo com o que foi apresentado na Tabela 4, as distribuições dos segmentos são apresentados em tabelas classificação cruzada e, seguindo a suposição de independência condicional, são identificadas as probabilidades de se observar o pertencimento de cada resposta a cada classe latente por meio da estimação dos parâmetros do modelo.

Pelo apresentado nessa tabela, é possível verificar de imediato as características gerais de cada um dos nove segmentos selecionados. Consequentemente, o perfil do estudante da Universidade de Brasília pode ser descrito por estes nove segmentos a seguir.

- **Segmento 1:** É composto por estudantes majoritariamente do sexo feminino e predominantemente amarelos/brancos que ingressaram pelo sistema de cotas sociais sem a realização de curso preparatório, que estudaram em escola pública no médio, e os pais desses estudantes possuem escolaridade com ensino médio completo. A renda mensal familiar encontra-se nas categorias mais baixas de zero a dez salários mínimos.
- **Segmento 2:** É formado por estudantes majoritariamente amarelos/brancos e do sexo masculino, que estudaram em escola particular e ingressaram pelo sistema universal sem a realização de curso preparatório. Os pais desses estudantes possuem escolaridade com ensino superior completo e a renda mensal familiar está disposta na categoria mais alta, acima de vinte salários mínimos.
- **Segmento 3:** É constituído por estudantes que ingressaram pelo sistema de cotas raciais sem a realização de curso preparatório e são provenientes de escolas públicas. São pretos/pardos, majoritariamente do sexo feminino, mas também com pertencimento relevante ao sexo masculino. Os pais possuem escolaridade com ensino médio completo e a renda familiar mensal predominante é de três a dez salários mínimos.
- **Segmento 4:** É integrado por estudantes que ingressaram pelo sistema universal sem curso preparatório com ensino médio cursado em escola pública. São na maioria amarelos/brancos, mas também com pertencimento relevante à classe dos pretos/pardos, e majoritariamente do sexo feminino, mas também com pertencimento relevante ao sexo masculino. Os pais possuem baixa escolaridade de terem cursado no máximo até o ensino fundamental e a renda familiar mensal encontra-se na classificação mais baixa, de zero até três salários mínimos.
- **Segmento 5:** É configurado por estudantes pretos/pardos, que ingressaram pelo sistema de cotas raciais sem curso preparatório e procedentes de um ensino médio realizado em escola pública. São majoritariamente do sexo feminino, mas também com pertencimento relevante ao sexo masculino. Os pais possuem baixa escolaridade, cursaram até o ensino fundamental, e possuem baixa renda mensal familiar, de até três salários mínimos.
- **Segmento 6:** É representado por estudantes que amarelos/brancos que ingressaram pelo sistema de cotas sociais, oriundos de escola pública, sem a realização de curso preparatório. Com pertencimento pareado em relação ao sexo, podem ser do sexo masculino ou feminino, seus pais possuem baixa escolaridade, cursaram até o ensino fundamental, e a renda mensal familiar do estudante é a mais baixa, de até três salários mínimos.

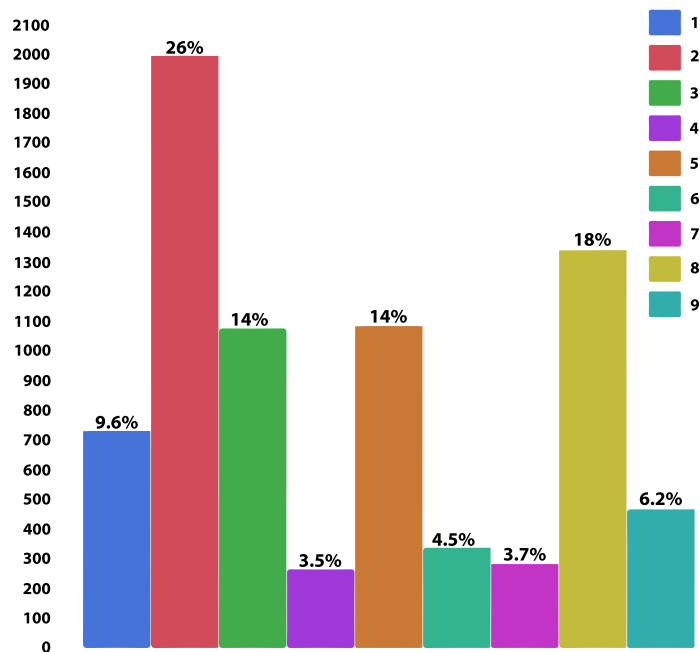
- **Segmento 7:** É formado por estudantes que ingressaram pelo sistema de cotas sociais e vieram de escola pública sem a realização de curso preparatório. São majoritariamente do sexo masculino, amarelos/brancos e os pais possuem escolaridade com ensino superior completo. A renda mensal familiar da maioria é de três a dez salários mínimos, mas também possui pertencimento à categoria de dez a vinte salários mínimos.
- **Segmento 8:** É constituído por estudantes que ingressaram através do sistema universal sem a realização de curso preparatório e vieram de escola particular. São majoritariamente amarelos/brancos, do sexo feminino, com pai de escolaridade com ensino médio completo e mãe com ensino superior completo e a renda mensal familiar é de três a dez salários mínimos.
- **Segmento 9:** É composto por estudantes pretos/pardos que ingressaram através do sistema universal com a realização de curso preparatório e estudaram em escola particular. São da maioria do sexo masculino, seus pais possuem o ensino superior completo e a renda mensal familiar encontra-se na faixa de três a dez salários mínimos.

De acordo com as probabilidades de pertencimento expressas na Tabela 4, constatou-se que a variável sexo não apresentou uma forte influência na segmentação dos estudantes. E que, na maioria dos casos, o grau de pertencimento era muito próximo entre as categorias logo a possibilidade de indivíduos de ambos os sexos pertencerem ao segmento é existente.

Tabela 5 – Frequências absolutas e relativas por segmentos

| S     | N    | %     |
|-------|------|-------|
| 1     | 730  | 9.6%  |
| 2     | 1992 | 26.3% |
| 3     | 1076 | 14.2% |
| 4     | 265  | 3.5%  |
| 5     | 1083 | 14.3% |
| 6     | 338  | 4.5%  |
| 7     | 282  | 3.7%  |
| 8     | 1338 | 17.7% |
| 9     | 467  | 6.2%  |
| Total | 7571 | 100%  |

Figura 9 – Percentual de ingressantes na UnB por segmento, 2017



Analisando a Tabela 5 e o gráfico apresentado na Figura 9, verifica-se que o segmento com o maior número de estudantes da Universidade de Brasília com características em comum encontra-se no segmento 2 (26%), seguido do segmento 8 (18%). Já a minoria encontra-se nos segmentos 4 (3.5%) e 7 (3.7%).

Tabela 6 – Frequências absolutas e relativas por segmentos em cursos de alto prestígio

| S     | N    | %     |
|-------|------|-------|
| 1     | 189  | 9.5%  |
| 2     | 667  | 33.6% |
| 3     | 270  | 13.6% |
| 4     | 26   | 1.3%  |
| 5     | 246  | 12.4% |
| 6     | 78   | 3.9%  |
| 7     | 115  | 5.8%  |
| 8     | 246  | 12.4% |
| 9     | 146  | 7.4%  |
| Total | 1983 | 100%  |

Tabela 7 – Frequências absolutas e relativas por segmentos em cursos de médio prestígio

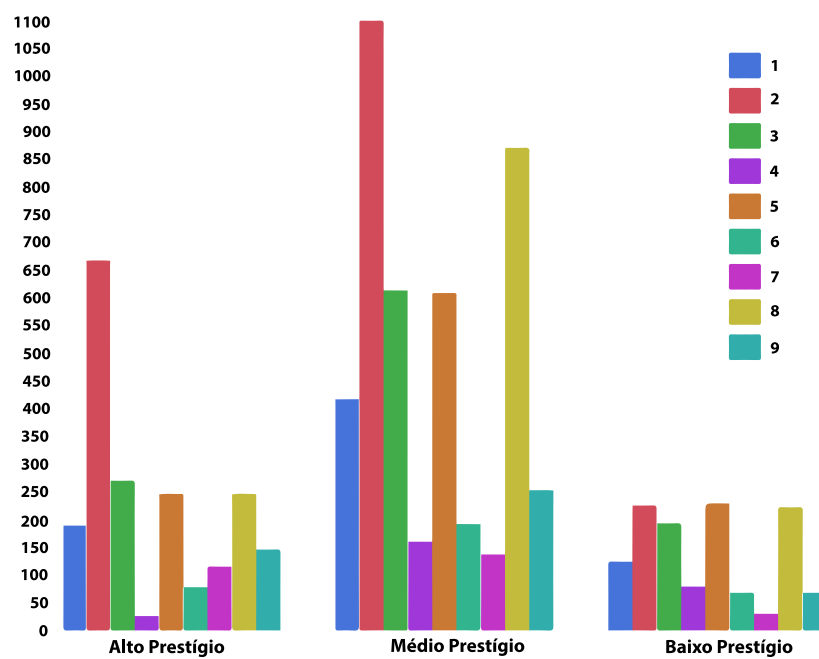
| S     | N    | %     |
|-------|------|-------|
| 1     | 417  | 9.6%  |
| 2     | 1099 | 25.3% |
| 3     | 613  | 14.1% |
| 4     | 160  | 3.7%  |
| 5     | 608  | 14.0% |
| 6     | 192  | 4.4%  |
| 7     | 137  | 3.2%  |
| 8     | 870  | 20.0% |
| 9     | 253  | 5.8%  |
| Total | 4349 | 100%  |

Tabela 8 – Frequências absolutas e relativas por segmentos em cursos de baixo prestígio

| S     | N    | %     |
|-------|------|-------|
| 1     | 124  | 10.0% |
| 2     | 226  | 18.2% |
| 3     | 193  | 15.6% |
| 4     | 79   | 6.4%  |
| 5     | 229  | 18.5% |
| 6     | 68   | 5.5%  |
| 7     | 30   | 2.4%  |
| 8     | 222  | 17.9% |
| 9     | 68   | 5.5%  |
| Total | 1239 | 100%  |

Quando comparados os segmentos obtidos por o prestígio do curso, verifica-se que os estudantes pertencentes ao segmento 2 estão presentes entre os de maior concentração mesmo em classes diferentes de prestígio do curso. Sendo que, no alto e no médio prestígio, com 33.6% e 25.3% respectivamente, lideram com unanimidade, seguido pelo segmentos 3 (13.6%), em cursos de alto prestígio, e 8 (20%) em cursos de médio prestígio. Em cursos de baixo prestígio a maioria encontra-se no segmento 5 (com 18.5%) seguido pelo segmento 2 (18.2%) e do segmento 8 (17.9%).

Figura 10 – Número de ingressantes na UnB segundo prestígio do curso por segmento, 2017



## 8 Conclusão

Com o objetivo de identificar o perfil do estudante da Universidade de Brasília, conclui-se que, no geral, o segmento mais homogêneo e de maior concentração evidenciado em relação aos nove grupos foi o de estudantes majoritariamente amarelos/brancos, que estudaram em escola particular e ingressaram pelo sistema universal sem a realização de curso preparatório. Os pais desses estudantes possuem escolaridade com ensino superior completo e a renda mensal familiar está disposta na categoria mais alta, acima de vinte salários mínimos. Essas características são notabilizadas pelo segmento 2 e representam 26.3% de todos os estudantes da Universidade de Brasília.

Verificou-se também, que independente do prestígio do curso, esse perfil de estudante se faz presente com características mais definidas e entre as maiores proporções dentre os segmentos. Além disso, constatou-se que, quanto mais baixo é o prestígio do curso, menor é essa desigualdade até que, para os cursos de baixo prestígio, tem-se as proporções dos segmentos de forma mais pareia e um perfil mais diverso do estudante ingressante na UnB. Esse fenômeno acontece com a maioria dos segmentos e pode ser evidenciado mais claramente ao observar os segmentos 3 e 5 que passam a ter maior participação em relação ao determinado prestígio do curso. Assim como o 8, que aparece com maior relevância nos cursos de médio e baixo prestígio.

Porém, assim como o segmento 2, outros nem tão explícitos tendem a decrescer conforme o prestígio do curso diminui. Esse é o caso dos segmentos 7 (estudantes que ingressaram pelo sistema de cotas sociais e vieram de escola pública sem a realização de curso preparatório, amarelos/brancos e os pais possuem escolaridade com ensino superior completo. A renda mensal familiar da maioria é de três a dez salários mínimos, mas também possui pertencimento à categoria de dez a vinte salários mínimos), e do 9 (estudantes pretos/pardos que ingressaram através do sistema universal com a realização de curso preparatório e estudaram em escola particular. Seus pais possuem o ensino superior completo e a renda mensal familiar encontra-se na faixa de três a dez salários mínimos).

Além disso, vale destacar a baixa inserção do segmento 4, principalmente em cursos de alto prestígio. Esse segmento constitui um perfil de estudante que ingressou pelo sistema universal sem curso preparatório com ensino médio cursado em escola pública, podem ser amarelos/brancos ou pretos/pardos. Os pais possuem baixa escolaridade de terem cursado no máximo até o ensino fundamental e a renda familiar mensal encontra-se na classificação mais baixa, de zero até três salários mínimos.)

Portanto, constatou-se que, a escolaridade dos pais e a renda familiar são fatores de grande influência para a identificação de tipologias dos ingressantes de graduação

na Universidade de Brasília, ainda mais quando relacionados com o prestígio do curso escolhido.



# Referências

AGRESTI, A. *Categorical data analysis*. New York [u.a.]: Wiley, 1990. XV, 558 S. p. (A Wiley-Interscience publication). Nenhuma citação no texto.

AKAIKE, H. *Information Theory and an Extension of the Maximum Likelihood Principle*. New York, NY: Springer New York, 1973. 199–213 p. Nenhuma citação no texto.

CALADO, C. da E. *Modelos de mistura em CRM: uma aplicação à segmentação no sector bancário*. Dissertação (Mestrado) — Lisboa: ISCTE, 2008. Nenhuma citação no texto.

CASTRO, M. H. G. *Sistemas nacionais de avaliação e de informações educacionais*. [S.l.]: Perspectiva, 2000. Nenhuma citação no texto.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the *EM*, algorithm. *Journal of the Royal Statistical Society*, 1977. Nenhuma citação no texto.

DILLON, B.; MUKHERJEE, S. *A guide to the design and execution of segmentation studies*. In R. Grover M. Vriens *The handbook of marketing research: Uses, misuses, and future advances* (pp. 523-545). [S.l.]: SAGE Publications Ltd, 2006. Nenhuma citação no texto.

FECHINE, V. M. R. *Um método de Segmentação Utilizando o Teste Qui-Quadrado para Variáveis Categorizadas*. Dissertação (Mestrado) — Universidade de Brasília - UnB, 1999. Nenhuma citação no texto.

FRANK, R. E.; MASSY, W. F.; WIND, Y. *Market segmentation*. [S.l.]: Prentice-Hall, 1972. Nenhuma citação no texto.

KAMAKURA; A, D. W.; MAZZON, J. A. *Estratificação Socioeconômica e consumo no Brasil*. [S.l.]: Editora Blucher - São Paulo, 2013. Nenhuma citação no texto.

KARINO, C. A. *Avaliação da igualdade, equidade e eficácia no sistema educacional brasileiro*. Tese (Doutorado) — Universidade de Brasília - UnB, 2016. Nenhuma citação no texto.

KOTLER, P. *Marketing Management*. [S.l.]: Prentice-Hall, 1988. Nenhuma citação no texto.

LANZA, S. T. et al. *Proc lca proc lta: users' guide* (version 1.3.2). University Park: The Methodology Center - Penn State, 2005. Nenhuma citação no texto.

LINZER, D. A.; LEWIS, J. B. *polCA: An R package for polytomous variable latent class analysis*. *Journal of Statistical Software*, v. 42, n. 10, p. 1–29, 2011. Nenhuma citação no texto.

MCLACHLAN, G.; KRISHNAN, T. *The EM Algorithm and Extensions*. New York: Wiley, 1997. Nenhuma citação no texto.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v. 6, p. 461–464, 1978. Nenhuma citação no texto.

SILVA, J. S.; BARBOSA, J. L.; SOUSA, A. I. *Desigualdade e diferença na universidade: gênero, etnia e grupos sociais populares*. [S.l.]: Rio de Janeiro: UFRJ, Pró-Reitoria de Extensão, 2006. Nenhuma citação no texto.

WEDEL, M.; KAMAKURA, W. A. *Marketing Segmentation: conceptual and methodological foundations*. [S.l.]: Kluwer, 1999. Nenhuma citação no texto.

WIND, Y. *Issues and Advances in Segmentation Research*. [S.l.]: Markus Wiener, Incorporated, 1978. Nenhuma citação no texto.

YANG, M.; YU, N. Estimation of parameters in latent class models using fuzzy clustering algorithms. *European Journal of Operational Research*, v. 160, n. 2, p. 515–531, 2005. Nenhuma citação no texto.

# Anexos

# ANEXO A – Sintaxe de Programação SAS

listings

```
PROC FREQ DATA=DADOS;
TABLE V003C;
TABLE V008C;
TABLE V014C;
TABLE V036C;
TABLE V028C;
TABLE V029C;
TABLE V041C;
TABLE V024C1;
TABLE ABEP;
TABLE PRESTIGIO;
RUN;
PROC LCA DATA=DADOS;
NCLASS 2;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 3;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 4;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
```

```
RUN;
PROC LCA DATA=DADOS;
NCLASS 5;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 6;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 7;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 8;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 9;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
```

```
NCLASS 10;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 11;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;
PROC LCA DATA=DADOS;
NCLASS 12;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;

PROC LCA DATA=DADOS OUTPARAM=param9CM OUTPOST=post9CM OUTEST=est9CM
OUTSTDERR=stderr9CM;
NCLASS 9;
ID ID;
ITEMS V003C V008C V014C V036C V028C V029C V041C V024C1;
CATEGORIES 3 2 3 2 3 3 2 4;
RHO PRIOR = 1;
SEED 861551;
RUN;

PROC SORT DATA=DADOS;
BY ID;
RUN;
DATA DADOSPRESTIGIOCM;
SET DADOS (KEEP = ID PRESTIGIO ABEP);
RUN;
```

```
PROC SORT DATA=DADOSPRESTIGIOCM;  
BY ID;  
RUN;  
PROC SORT DATA=POST9CM;  
BY ID;  
RUN;  
DATA SEGMENTOSCM;  
MERGE DADOSPRESTIGIOCM POST9CM;  
BY ID;  
RUN;  
  
PROC FREQ DATA=SEGMENTOSCM;  
TABLE V003C*BEST;  
TABLE V008C*BEST;  
TABLE V014C*BEST;  
TABLE V036C*BEST;  
TABLE V028C*BEST;  
TABLE V029C*BEST;  
TABLE V041C*BEST;  
TABLE V024C1*BEST;  
TABLE ABEP*BEST;  
TABLE PRESTIGIO*BEST;  
RUN;
```